

# PLEM: a Web 2.0 driven Long Tail aggregator and filter for e-learning

Citation for published version (APA):

Chatti, M., Anggraeni, Jarke, M., Specht, M., & Maillet, K. (2010). PLEM: a Web 2.0 driven Long Tail aggregator and filter for e-learning. *International Journal of Web Information Systems*, 6(1), 5-23.  
<https://doi.org/10.1108/17440081011034466>

**DOI:**

[10.1108/17440081011034466](https://doi.org/10.1108/17440081011034466)

**Document status and date:**

Published: 01/01/2010

**Document Version:**

Peer reviewed version

**Document license:**

CC BY-NC-SA

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05 May. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



Journal TITLE:	
<b>International Journal of Web Information Systems</b>	
Article Title:	
<b>PLEM: A Web 2.0 Driven Long Tail Aggregator and Filter for e-Learning</b>	
Authors	
<b>Mohamed Amine Chatti, Anggraeni, Matthias Jarke, Marcus Specht and Katherine Maillet</b>	

## Abstract

The **Personal Learning Environment** (PLE) driven approach to learning suggests a shift in emphasis from a teacher driven knowledge-push to a learner driven knowledge-pull learning model. One concern with knowledge-pull approaches is knowledge overload. The concepts of **collective intelligence** and **the Long Tail** provide a potential solution to help learners cope with the problem of knowledge overload. Based on these concepts, the paper proposes a filtering mechanism that taps the collective intelligence to help learners find quality in the Long Tail, thus overcoming the problem of knowledge overload. We present theoretical, design, and implementation details of PLEM, a Web 2.0 driven service for personal learning management, which acts as a Long Tail aggregator and filter for learning. The primary aim of PLEM is to harness the collective intelligence and leverage social filtering methods to rank and recommend learning entities.

## Keywords

Personalization, Personal Learning Environment, Web 2.0, e-Learning 2.0, The Long Tail, Collective Intelligence, Wisdom of Crowds, Social Filtering

## **MAIN CONTENTS SUBMISSION**

<b>PLEM: A Web 2.0 Driven Long Tail Aggregator and Filter for e-Learning</b>

**Abstract**

The **Personal Learning Environment** (PLE) driven approach to learning suggests a shift in emphasis from a teacher driven knowledge-push to a learner driven knowledge-pull learning model. One concern with knowledge-pull approaches is knowledge overload. The concepts of **collective intelligence** and **the Long Tail** provide a potential solution to help learners cope with the problem of knowledge overload. Based on these concepts, the paper proposes a filtering mechanism that taps the collective intelligence to help learners find quality in the Long Tail, thus overcoming the problem of knowledge overload. We present theoretical, design, and implementation details of PLEM, a Web 2.0 driven service for personal learning management, which acts as a Long Tail aggregator and filter for learning. The primary aim of PLEM is to harness the collective intelligence and leverage social filtering methods to rank and recommend learning entities.

**Keywords**

Personalization, Personal Learning Environment, Web 2.0, e-Learning 2.0, The Long Tail, Collective Intelligence, Wisdom of Crowds, Social Filtering

**1. Introduction**

There is a wide agreement that the new era of education is defined by rapid knowledge development. Brown and Adler (2008), for instance, note: "In the twentieth century, the dominant approach to education focused on helping students to build stocks of knowledge and cognitive skills that could be deployed later in appropriate situations. This approach to education worked well in a relatively stable, slowly changing world in which careers typically lasted a lifetime. But the twenty-first century is quite different. The world is evolving at an increasing pace". It is suggested that self-organized learning (also known as self-directed or self-determined learning) is appropriate to the needs of learners in the twenty-first century, particularly in the development of individual capability (Hase & Kenyon, 2000).

In recent years, the concept of Personal Learning Environment (PLE) has been widely discussed among Technology Enhanced Learning (TEL) researchers, as a natural and learner-centric model that supports the self-organized learning process by surrounding the learner with the environment that matches her needs best. One concern with a PLE driven approach to learning is knowledge overload. Over the past years, Web 2.0 technologies have offered abundant access to knowledge. A massive choice could be however overwhelming for learners. Collective intelligence, when the whole is greater than the sum of its

parts, can help learners cope with the problem of knowledge overload. And, Long Tail aggregators and filters, driven by collective intelligence, can provide a potential solution to overcome the problem of knowledge overload. This work builds on these two important concepts, and presents theoretical, design, and implementation details of PLEM as a Web 2.0 driven service for personal learning management, which acts as a Long Tail aggregator and filter for learning. The primary aim of PLEM is to harness the collective intelligence and leverage social filtering methods to help learners locate quality learning entities.

The idea behind the PLEM aggregation and filtering mechanism is quite simple. Each distributed filtering action on a learning element from the Web (e.g. comment, link, save, like, rate, vote, view, share) counts as one "vote" for that learning element. The popularity of a learning element is then measured by aggregating the number of "votes" for that learning element, gathered from multiple distributed Web 2.0 services.

The paper proceeds as follows. In section 2, we introduce theoretical background of PLEs and relevant design issues for personal environments. In sections 3 and 4, we describe the associated problem of knowledge overload and how concepts of collective intelligence can help to overcome this problem. In section 5, we highlight the role of knowledge aggregation and filtering for effective Long Tail support in learning environments. We follow in section 6 with the design and implementation of PLEM as a personal learning environment that supports self-directed learners with knowledge aggregation and filtering. And finally, we summarize our findings in section 7.

## **2. Personal Learning Environments**

Among others, Hase & Kenyon (2000) argue that the rapid rate of change in society suggests that we should now be looking at a learning approach where it is the learner himself who determines what and how learning should take place, and point out that self-organized learning may well provide the optimal approach to learning in the twenty-first century.

Self-organized learning provides a base for the establishment of a model of learning that goes beyond curriculum and organization centric models, and envisions a new learning model characterized by the convergence of lifelong, informal, and ecological learning within a learner-controlled space. Knowles (1970, p7; cited in Hase & Kenyon, 2000) defined self-organized learning as: "The process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing learning strategies, and evaluating learning outcomes". The idea of self-organized learning has been put forward by the concept of double-loop learning introduced by Argyris and Schön (1978) within an organizational setting. Argyris (1991) argues that most people define learning too narrowly as mere "problem solving", so they focus on identifying and correcting errors in the external environment. This is what Argyris calls single-loop learning. But, in

the words of Argyris:

*If learning is to persist, managers and employees must also look inward. They need to reflect critically on their own behavior, identify the ways they often inadvertently contribute to the organization's problems, and then change how they act.*

This deeper form of learning is what Argyris terms "double-loop learning". Argyris and Schön (1996, p20) define single-loop learning as "learning that changes strategies of actions or assumptions underlying strategies in ways that leave the values of a theory of action unchanged", and double-loop learning as "learning that results in a change in the values of theory-in-use, as well as in its strategies and assumptions". In other words, Argyris and Schön differentiate between learning that does not change the underlying mental models of the learners but merely revises their application scenarios (single-loop), and learning which does affect such changes (double-loop). Double-loop learning starts from a learner's mental model (i.e. theories-in-use) defined by base norms, values, strategies, and assumptions, and suggests critical reflection in order to challenge, invalidate, or confirm the used theories-of-use. The result of this reflection would be a reframing of one's norms and values, and a restructuring of one's strategies and assumptions, according to the new settings. Double-loop learning, thus, requires self-criticism, i.e. the capacity for questioning one's theories-in-use and encourages inquiry into and testing of one's actions.

In recent years self-organized learning is increasingly supported by responsive, open, and personal learning environments, where the learner is in control of her own development and learning. The Personal Learning Environment (PLE) concept translates the principles of self-organized learning into actual practice. PLE is a relatively new term, first introduced in 2004 (van Harmelen, 2006). van Harmelen (ibid.) describes PLEs as:

*Systems that help learners take control of and manage their own learning. This includes providing support for learners to*

- set their own learning goals*
  - manage their learning; managing both content and process*
  - communicate with others in the process of learning*
- and thereby achieve learning goals.*

A PLE-driven approach to learning gets beyond centralized learning management systems and supports a wide variety of learning experiences outside the institutional boundaries. A PLE suggests the freeform use of a set of lightweight and loosely coupled tools and services that belong to and are controlled by individual learners. Rather than being restricted to a limited set of services within a centralized institution-controlled system, the idea is to provide the learner with a plethora of different services and hand over control to her to select, use, and remix the services the way she deems fit. A PLE does not only provide personal spaces, which belong to and are controlled by the learner, but also requires a social context by offering means to connect with other personal spaces for effective knowledge sharing and collaborative knowledge creation.

A PLE-driven approach to learning also suggests a shift in emphasis from a knowledge-push to a knowledge-pull learning model. In a learning model based on knowledge-push, the information flow is directed by the institution/teacher. In a learning model driven by knowledge-pull, the learner navigates toward knowledge. One concern with knowledge-pull approaches is knowledge overload.

### **3. Dealing with Knowledge Overload**

Information and communication technology has made information abundant and easily accessible. New information comes every day, forms a knowledge rich world and at the same time shortens the lifetime of the information itself. In this world of unlimited space and abundance, people are increasingly brought into near limitless choices of almost everything, making them suffer from increasing complexity and information overload. Heylighen (2002) discusses the effects of the information overload phenomenon. The author notes that the longer people are subjected to information overload, the more negative its effects on physical and mental well-being. The result of information overload is that well-being is replaced by anxiety, loss of control, and stress. The author further points out that information overload also leads to both "subjective frustration, where people feel anxious or guilty because they think they may have missed essential elements, and objective failure, where wrong decisions are made because not enough information was taken into account". Personalization and information filtering is often described as one possible solution for the problem information overload. One approach to personalization is building on collective intelligence.

Collective intelligence can play an essential role to cope with problems caused by knowledge overload. Collective intelligence is a name for the synergetic use of individually intelligent components (Levy, 1997). It makes that even complex behavior may be coordinated by relatively simple interaction (Miller, 2007). Ant colonies provide a good example of complex systems driven by collective intelligence, in which the parts use only local information and the whole thing directs itself. Operating as a collective, an ant colony can solve problems that would be unthinkable for individual ants, such as finding the shortest path to the best food source, allocating workers to different tasks, or defending a territory from neighbors. As Deborah Gordon, a biologist at Stanford University studying harvester ants in the Arizona desert, puts it: "Ants aren't smart. Ant colonies are". The coordinated behavior of ant colonies arises from the ways that ants use local information. Take foraging as an example. In ant colonies, patrollers use chemical trails (pheromones) to lead foragers to food resources. Foragers then tend to leave in the direction that the patrollers return from. As more foragers use the same path to food resources, the pheromone trail gets reinforced and the path becomes more attractive to fellow foragers (Miller, 2007). Gordon (2007) gives another example illustrating the same concept: "A forager won't come back until it finds something. The less food there is, the longer it takes the forager to find it and get back. The more food there is, the

faster it comes back. So nobody's deciding whether it's a good day to forage. The collective is, but no particular ant is".

James Surowiecki (2004) took this thinking further with his book: *The Wisdom of Crowds*. In the introduction of his book, Surowiecki notes that "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them" and that "group's decisions will, over time, be intellectually superior to the isolated individual, no matter how smart or well-informed he is". A crowd, Surowiecki argues, is "more than just the sum of its members. Instead, it was a kind of independent organism. It had an identity and a will of its own, and it often acted in ways that no one within the crowd intended". The author, however, stresses that not all crowds are smart. There are conditions that are necessary for the crowd to be wise, namely diversity of opinion, independence, decentralization, and aggregation.

According to Surowiecki, a group, which includes members with wide diversity of knowledge, ability or skill has better chance to come up with good group decision. It happens because, as a human being, each group member might only have a fraction of everything she needs to know. An individual only owns private and limited information. No matter how valuable and accurate the information she has, it is still partial and incomplete. In order to produce near-ideal decision, a group needs to widen its perspectives. Diversity adds perspective that would otherwise be absent. It allows the group to measure the problem from different angles.

Diversity of opinion leads to independence, another key criteria of good group decision-making. As Surowiecki puts it: "a group of people is far more likely to come up with a good decision if the people in the group are independent of each other" (p. 41). Surowiecki further points out that "independence is important to intelligent decision making for two reasons. First, it keeps the mistakes that people make from becoming correlated. Errors in individual judgment won't wreck the group's collective judgment as long as those errors aren't systematically pointing in the same direction...Second, independent individuals are more likely to have new information rather than the same old data everyone is already familiar with" (p. 41)

The idea of wisdom of crowd also takes decentralization as an important criterion. Decentralization implies that "if you set a crowd of self-interested, independent people to work in a decenaralized way on the same problem, instead of trying to direct their efforts from the top down, their collective solution is likely to be better than any other solution you could come up with" (p. 70). Decentralization is important because it increases the scope and diversity of the opinions and information in the system. Surowiecki stresses that, to solve a problem, individuals should rely on their local and specific knowledge rather than rely on an omniscient or farseeing planner. Decentralization is also crucial when dealing with tacit knowledge. At the heart of decentralization is the assumption that "the closer a person is to a problem, the more likely he or she is to have a good solution to it" (p. 71). One success



story of decentralization, mentioned in Surowiecki's book, is the discovery of the SARS virus. It includes collaboration work of eleven research laboratories around the world. According to Surowiecki, the intriguing thing about the success is that the laboratories work in decentralization way to find the cause of the SARS disease. No one at the top dictating what each lab should do, what they should work on, how they exchanged information. Strictly speaking, no one was in charge of it.

However, decentralization itself is not enough. Since each individual works locally, it is possible that her valuable information cannot be noticed by the rest of the system. In order to make individual knowledge collectively useful, a system needs a way to aggregate all local and private information into a collective whole. As Surowiecki writes: "a balance between the local and the global is essential: a decentralized system can only produce genuinely intelligent results if there's a means of aggregating the information of everyone in the system. Without such a means, there's no reason to think that decentralization will produce a smart result" (p. 74). Aggregation is thus important to the success of decentralization. Surowiecki summarizes this idea noting, "Centralization is not the answer. But aggregation is" (p. 78).

#### **4. Collective Intelligence in Web 2.0**

The term Web 2.0 has been used to describe a social -not technological - evolution of the Web from being a medium, in which information was transmitted and consumed, into being a platform, in which content was created, shared, remixed, repurposed, and passed along (Downes, 2005).

Harnessing collective intelligence has become the driving force behind Web 2.0. As O'Reilly (2007) puts it: "Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. This is what I've elsewhere called "harnessing collective intelligence"".

The Internet has allowed people from all over the world to be connected with each other in a way that was never imagined before. Its latest evolution turns the Web into a global community where people can work and collaborate in a new way. It is important to take advantage of this phenomenon. The challenge now is to make people who broadly connected via the Internet acting smarter than any individuals can do. Many Web 2.0 applications embrace user's contributions to amplify their values. For instance, Google's PageRank uses the collective intelligence of the Web to determine a page's importance. It uses the link structure of the Web as weighted votes to decide which page contains the most useful and relevant information. Amazon always leads its search with most popular items based on the sales and other customer flow activity around similar products. Wikipedia, the open encyclopedia, created not by paid experts and editors, but by whoever wants to contribute, shows that collectively we

know far more than one single person does. Moreover, different sites and products on the Web rely on “viral marketing” from one user to another rather than from advertisement to get exposed. YouTube’s rating scheme, eBay’s feedback, Digg’s voting are also successful attempts to harness user’s collective intelligence on the Web (O’Reilly, 2007). Social bookmarking, social tagging, and folksonomies are also successful examples of the collective intelligence in action, as users share, organize, filter interesting information for each other, browse related topics, discover unexpected resources that otherwise they would never know existing, look for what others have tagged, subscribe to an interesting tag and receive new content labeled with that tag via Web feeds, and find unknown people with similar interests (Chatti & Jarke, 2009).

## **5. The Long Tail**

The theory of the Long Tail presents a framework to synthesize the results and concepts discussed in the last two sections.

### **5.1 The Theory of the Long Tail**

Chris Anderson (2008) in his book “The Long Tail” describes a phenomenon about how the focus of our market and culture is shifting from only few popular items into millions of otherwise niche items. As he puts it, “increasingly, the mass market is turning into a mass of niches” (p. 5). This phenomenon happened not because people have changed their interests. It happened because “technology is turning mass markets into millions of niches” (p. 15).

Figure 1 illustrates a power law graph, which is commonly included within the standard discussion of The Long Tail. It demonstrates ranking of popularity. The horizontal axis represents items and services that are available for sale. The vertical ordinate shows number of units sold for each item. To the left of vertical line is the Short Head, mainstream items, which dominate the sale. To the right is the Long Tail, niche items, which are usually ignored by many traditional stores as consequence of limited storage shelf. Anderson notes that in traditional store with lack of storage shelf, only the Short Head items are available for sale. In the era of unlimited storage space, however, niche products can be as economically attractive as mainstream products. In statistics, curves like the one shown in Figure 1 are called “long-tailed distributions”, because the tail of the curve is very long relative to the head. Anderson borrows the term to describe his theory of the Long Tail as follows: “Our culture and economy are increasingly shifting away from a focus on a relatively small number of hits (mainstream products and markets) at the head of the demand curve, and moving toward a huge number of niches in the tail” (p. 52). He further points out that “our culture is increasingly a mix of head and tail, hits and niches, institutions and individuals, professionals and amateurs. Mass culture will not fall, it will simply get less mass. And niche culture will get less obscure” (p. 182).

Anderson suggests six themes characterizing the Long Tail: (1) Far more niche goods than hits, (2) Costs of reaching niches is falling dramatically, (3) Filters can drive demand down the Tail, (4) The demand curve flattens, (5) Many niche products are a market as big as the hit market, (6) Natural shape of demand is revealed. Anderson further identifies three forces representing a new set of opportunities in Long Tail to reduce the costs of reaching niches.

The first force is democratizing the tools of production. Cheap and ubiquitous digital technologies of production and tools of creativity caused that the traditional line between professional and amateur producers have blurred. Everybody has now a better opportunity to take active role on fields of interest and a better chance to find a real audience. The result, Surowiecki notes, is that "the available universe of content is now growing faster than ever. That is what extends the tail to the right, increasing the population of available goods manyfold" (p. 54). The democratized tools of production are thus leading to the problem of knowledge overload that we discussed in Section 3.

The second force is democratizing distribution. As the tail of niche products is currently getting longer, people need something that can connect them with the products. "Aggregators" are a manifestation of the second force. According to Anderson, aggregators "lower the barrier to market entry, allowing more and more things to cross that bar and get out there to find their audience" (p. 88). He defines a Long Tail aggregator as "a company or service that collects a huge variety of goods and make them available and easy to find, typically in a single place" (p. 88), and stresses that "successful Long Tail aggregators need to have both hits and niches" (p. 148). Anderson also gives several prime examples of popular Long Tail aggregators. For instance, Google aggregates the Long Tail of advertising. iTunes aggregates the Long Tail of music. Netflix does the same for the Long Tail of movies. eBay aggregates the Long Tail of physical goods and the Long Tail of people who sell them. Feed readers order the Long Tail of online content, including millions of blogs. And, Wikipedia is an aggregator of the Long Tail of knowledge and those who have it.

The third force is connecting supply and demand, "introducing consumers to new and newly available goods and driving demand down the tail, from hits to niches" (p. 55). The third force "increases demand for the niches and flattens the curve, shifting its center of gravity to the right" (p. 57). The two first forces; i.e. democratizing production and democratizing distribution, lead to exposing huge variety of items to people. On one hand, it is good because it gives people more choice and allows them to find what is right for them. On the other hand, more choice is sometimes not only confusing but also oppressive, especially when it contains so much irrelevant materials. It takes great effort to spot the quality among other unrelated random items. We therefore do need "filters" to screen out unqualified contents and focus more on suitable candidates. According to Anderson, "Amplified word of mouth is the manifestation of the third force of the Long Tail" (p. 107). This can take different forms such as recommendations, rating, ranking, reviews, comments, and votes. The effect of

Long Tail filters for consumers is to lower the search costs of finding niche content.

Anderson summarizes the three Long Tail forces as follows: "The first force, democratizing production, populates the Tail. The second force, democratizing distribution, makes it all available...[The] third force...helps people find what they want in this new superabundance of variety" (p. 107).

## **5.2 The Long Tail in Learning**

The theory of the Long Tail phenomenon can also be applied in the learning domain. As Brown & Adler (2008) point out: "Whereas traditional schools offer a finite number of courses of study, the "catalog" of subjects that can be learned online is almost unlimited. There are already several thousand sets of course materials and modules online, and more are being added regularly. Furthermore, for any topic that a student is passionate about, there is likely to be an online niche community of practice of others who share that passion".

Educational institutions have abandoned the Long Tail of content for decades. They used "hits" learning resources. Teachers were trained to teach contents which are claimed to be important for students - the "hits" in each subject area. Exams were designed to measure students' knowledge of these hits.

Web 2.0, on the other hand, offered unlimited opportunities for students to learn and explore subjects they love. A wide range of information and learning resources is now available, including the rarely used materials. Increasingly, educational institutions open access to their educational materials to anyone who wants to use them. This results in a rapidly growing amount of Open Educational Resources (OER). Examples include the MIT OpenCourseWare project ([ocw.mit.edu/](http://ocw.mit.edu/)), which opens the course materials (e.g. lecture notes, video lectures, problem sets and solutions, exams) that are used in the teaching of almost all of MIT's undergraduate and graduate subjects to any learner on the Web, free of charge. Another interesting move in the direction of OER is the recent launch of YouTube EDU ([YouTube.com/edu](http://YouTube.com/edu)), which centralizes the videos from over 100 universities and colleges, and gives access to free lectures, campus tours, research news, and academic content from leading universities, such as MIT, Stanford, and UC Berkeley. Moreover, Web 2.0 blurred the distinction between professional and amateur learners. The democratized tools of production in Web 2.0 are leading to a huge increase in the numbers of producers and a greater variety of learning resources. Infinite bandwidth, unlimited storage and easy tools to produce, edit and upload media through the Web has invited students to take active role. Students now have the chance to create their own content or mashup existing content to make new learning resources. This certainly populates the Long Tail of learning. Furthermore, Web 2.0 has also introduced a new participatory culture. Social networking services, for instance, allowed learners to go beyond the classroom and join a niche community of interest, where they can meet other learners with similar interests, share ideas, and collaborate in innovative ways.

The bottom line, Web 2.0 has offered abundant access to explicit knowledge nodes (i.e. learning resources) and tacit knowledge nodes (i.e. people). We learned from the theory of the Long Tail that massive choice could be overwhelming for learners. This is where Long Tail aggregators and filters can play a crucial role to foster learning. In the following sections, we present the details a PLEM, a Web 2.0 driven Long Tail aggregator and filter for learning, that taps the wisdom of crowds to help learners find quality in the Long Tail, thus enabling them to extend their PLEs with valuable knowledge nodes.

## 6. PLEM: Design and Implementation

In the ensuing sections, we will describe PLEM with an eye on the architectural and implementation details. The system design will be followed by a detailed description of the different modules and their underlying functionalities.

### 6.1 PLEM Design

An overview of the PLEM abstract architecture is provided in Figure 2. In PLEM, we distinguish between four types of learning entities, namely learning resource, learning service, learning expert, and learning community. The first two types represent explicit knowledge nodes and the last two represent tacit knowledge nodes. PLEM basically acts as a Long Tail aggregator and filter for learning. It aggregates niche learning entities within a single place and provides filter mechanisms that help learners identify appropriate learning entities. PLEM aggregation and filtering mechanisms are based on popular Web 2.0 concepts and social software technologies, such as social bookmarking, social tagging, folksonomies, OpenID, and mashups. OpenID is a free and easy way to use a single digital identity across the Internet, and is becoming the de facto standard Web protocol for user authentication. Mashup is another popular concept, often associated with Web 2.0. A mashup can be defined as the aggregation of different data sources and application programming interfaces (APIs) into an integrated web application. As a consequence of the The Web 2.0 movement, many service providers open their data sources and APIs to public or restricted communities. As a result, other parties try to use and combine the gathered data to come up with new services.

Figure 3 illustrates the detailed architecture of PLEM. PLEM is based on the Model-View-Controller (MVC) paradigm. The key idea behind the MVC pattern is a separation of concerns among the components responsible for the data (the model), the application logic (controller) and the web interface (view). The *model layer* in PLEM is a MySQL database, which provides persistent storage for learning entities. In the *view layer*, Google Web Toolkit (GWT) (Google, 2009) is used to generate the front-end in JavaScript and HTML. GWT is an open source Java development framework that allows developers to create AJAX applications in Java language using the Java development tools of their choice. The *controller layer* in PLEM is responsible for listening specific events that are defined in the view and handling all those events to the associated

model. Every interaction between the learner and the PLEM web interface is captured using Ajax script. The Ajax script captures the type of action chosen by the learner and determines the kind of request data it will send to the session handler. Some actions, e.g. creating, modifying, ranking, and tagging learning entities require authorization. In this case, learners should log in using their OpenID to be able to perform the actions. The session handler, then, will send the identity of the learner to the authentication module, which is responsible for the actual OpenID authentication. The database handler will manage all requests to the database. The management and aggregation module handles the management and aggregation of learning entities. The filtering module is responsible for ranking the learning entities, and is initiated in response to a learner search query. Both modules are based on mashups. They not only use data collected from PLEM users, but also data from third party service providers.

## **6.2 PLEM Implementation**

The following sections illustrate PLEM in action. In order to demonstrate how the system works, we show the functionalities of the different modules using actual examples. The focus will be on (1) the management and aggregation module and (2) the filtering module.

### **6.2.1 PLEM as a Long Tail Aggregator**

PLEM acts as a Long Tail aggregator for learning that collects a variety of niche learning entities (i.e. learning resources, learning services, learning experts, and learning communities) and makes them available and easy to find. A learner can log into PLEM and create a personalized space, where she can easily aggregate, manage, tag, rate, and share learning entities of interest. An example of such a space is depicted in Figure 4.

Today, learning resources are broken up into "microchunks" that are distributed over multiple domains. These microchunks are available in different forms such as texts, images, sounds, and videos. PLEM supports learners in reusing, remixing, and sharing learning resources with minimum effort. In PLEM, Learners can draw on a range of media to create their learning resource collections. This includes open educational resources provided by MIT OCW and OpenER (Open University Netherlands), blogposts, videos, books, images, and presentations.

The aggregation module in PLEM enables learners to pull together learning resources from more than one source, remix and assemble them to form a new learning resource collection. It provides a federated search engine (see Figure 5) that makes it possible to perform search across media and plug into multiple distributed domains to search for learning resources with a single query. This search engine is implemented as a mashup of different search engines and uses several open APIs provided by third party search services, such as Google Blog Search API and Technorati API for blogs, Google Video Search API and

YouTube API for videos, Google Book Search API for books, Google Image Search API and Flickr API for images, and Slideshare API for presentations. The learner can then gather the learning resources of her choice together into a personalized learning resource collection. Sequencing the newly aggregated learning resources is as simple as dragging and dropping them into the position the learner wants them within her learning resource collection. The self-compiled learning resource collection can be extended at a later time. It can also be shared with and reused by other learners. An example of an aggregated learning resource collection is shown in Figure 6.

#### **6.2.2 PLEM as a Long Tail Filter**

PLEM also acts as a Long Tail filter for learning. It taps the wisdom of crowds by following what learners do with learning entities and translating that into relevant search results. The PLEM filtering approach is an imitation of ant behavior. Learners are like ants searching for the best learning sources. Learners act as guides individually when they interact with learning entities on the Web (e.g. bookmark web pages, tag resources, recommend items, review books, comment on blogposts, trackback sites, share videos, vote on news), just as ants leave signals for other ants to show them the best trails. The idea is to leverage this distributed local filtering behavior to improve the search for relevant learning entities.

The PLEM filtering module uses a distributed voting mechanism to locate quality learning resources and services as well as appropriate communities and experts, based on a collective decision. Each filtering action on a learning entity (e.g. comment, link, save, like, rate, vote, view, share) counts as one vote for that learning entity. The mean value of all votings for a given learning entity is then used to measure its popularity.

The PLEM filtering module is thus built on the distributed intelligence of learners on the Web. It satisfies the four conditions that characterize wise crowds: diversity of opinion (each learner has some private information), independence (learner's opinions are not determined by the opinions of other learners), decentralization (learners act on local information), and aggregation (aggregation and filtering mechanisms exist for turning local individual judgments into a collective decision).

As shown in Figure 7, for each learning entity in PLEM, different interaction metrics are computed. These metrics currently include PLEM saves and ratings, Delicious saves, Friendfeed comments and likes, Yahoo inbound links, Digg votes, Google trackbacks, and Technorati blog reactions. This information is gathered using open APIs of the related services, and is used by the CoCoRank algorithm to rank learning entities. We present the details of the CoCoRank algorithm in the next section.

#### **6.2.3 The CoCoRank Algorithm**

The CoCoRank algorithm adopts both **content** and **conversation** mass metrics to

**rank** learning entities. The content mass metric checks the occurrence of a search term in the title, description, or keywords associated with the learning entities. The conversation mass metric is used to measure the popularity of a learning entity. This metric captures the mass of the total distributed conversation generated by a learning entity on the Web. The idea behind the CoCoRank algorithm is quite simple. We consider each simple interaction with a learning entity on the Web as a vote, and the learning entity that gets the most votes goes first on the list.

Figure 8 illustrates in an abstract manner how the algorithm works. PLEM provides the possibility to query its database based on tag, title and description, or creator (content mass). The list of learning entities, which satisfy the search query, is then ordered by popularity (conversation mass).

The CoCoRank algorithm assigns ranks to a set of learning entities that satisfy a user search query. Formally, the problem can be formulated as the following: Given a set of learning entities  $E = (e_1, e_2, \dots, e_n)$  which represents the learning entities that satisfy a user search query. Build a set of rank values  $R = (r_1, r_2, \dots, r_n)$  where  $r_i$  is the rank value of  $e_i$ .

The CoCoRank algorithm follows two steps and can be presented as following:

*Input:* a set of learning elements  $E = (e_1, e_2, \dots, e_n)$

*Output:* a set of rank value  $R = (r_1, r_2, \dots, r_n)$  where  $r_i$  is the rank value of  $e_i$ .

1. Collect votes (i.e. interaction metrics) on a learning element from various services.
2. Compute the rank value of the learning entity.

#### *Collect the votes*

In this step, we collect votes on a learning entity  $e$  from various services. Thereby, each interaction with the learning entity is considered as a vote. Having a set of services  $S = (s_1, s_2, \dots, s_m)$ , we send queries to the different services with the URL of the learning entity as a parameter. We then collect the votes on  $e$  coming from the different services in form of number of saves, entries, likes, comments, trackbacks, or inbound links. The output of this step is a set of vote vectors  $V$  in which  $v(e) = (v_1(e), v_2(e), \dots, v_m(e))$  represents the votes on the learning entity  $e$ ;  $v_j(e)$  is the number of votes on  $e$  retrieved from service  $s_j$ .

#### *Compute the rank value*

The rank value of a learning entity is determined by combining the votes retrieved from various services. The rank value of each service  $s$  for a learning entity  $e$  is equivalent to multiplying the number of votes on  $e$  that are retrieved from  $s$  and the prestige of  $s$ . It is defined as follows:

$$r_j(e) = p_j \cdot v_j(e),$$

where  $r_j$  is the rank value of service  $s_j$ ,  $p_j$  is the prestige of  $s_j$ , and  $v_j$  represents the number of votes retrieved from  $s_j$ .



The rank value for a learning entity  $e$  is then calculated as the sum of the rank values of all services in  $S$  for  $e$ . It is defined as follows:

$$r(e) = \sum_{j=1}^m r_j(e)$$

In the current version of the CoCoRank algorithm, we assign  $p_j = 1$  for each service  $s_j$  in  $S$ .

#### 6.2.4 Comparison of CoCoRank, PageRank and HITS

**PageRank** is a link analysis algorithm proposed by Brin & Page (1998) for identifying the importance of Web pages. The intuition behind PageRank is quite simple: A Web page is important if many important pages link to it. Thus, a link from a popular page is given a higher weighting than one from an unpopular page. Brin & Page (ibid.) define PageRank as follows:

*We assume page  $A$  has pages  $T1...Tn$  which point to it (i.e., are citations). The parameter  $d$  is a damping factor which can be set between 0 and 1. We usually set  $d$  to 0.85. There are more details about  $d$  in the next section. Also  $C(A)$  is defined as the number of links going out of page  $A$ . The PageRank of a page  $A$  is given as follows:*

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

PageRank can be calculated using an iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the Web. A major advantage of PageRank is that it is a global measure and is query independent. That is, the PageRank values of all pages on the Web are computed and saved off-line rather than at the query time. It is thus very efficient at query time (Liu, 2006).

Another popular link analysis algorithm is **HITS**, which stands for Hypertext Induced Topic Search (Kleinberg, 1999). HITS distinguishes between two types of Web pages. An **authority** is a page with many in-links. A **hub** is a page with many out-links. The key idea of HITS is that a good hub points to many good authorities and a good authority is pointed to by many good hubs. Thus, authorities and hubs have a mutual reinforcement relationship. HITS assigns every page an authority score and a hub score as follows: (Liu, 2006)

*Let the authority score of the page  $i$  be  $a(i)$ , and the hub score of page  $i$  be  $h(i)$ . The mutual reinforcing relationship of the two scores is represented as follows:*

$$a(i) = \sum_{(j,i) \in E} h(j)$$

$$h(i) = \sum_{(i,j) \in E} a(j)$$

Unlike PageRank, HITS is search query dependent. That is, when the user issues a search query, HITS first expands the list of relevant pages returned by a search engine and then produces the authority scores and the hub scores of the expanded set of pages (Liu, ibid.).

The main difference between CoCoRank and PageRank or HITS is that CoCoRank does not only use the link structure as an indicator of an individual page's value, but extends this to include other interactions with that page, such as comments, saves, likes, rates, votes, views, and shares. A detailed comparison of the three algorithms is provided in Table 1.

Table 1. Comparison of CoCoRank, PageRank and HITS

	PageRank	HITS	CoCoRank
Data set	Based on inbound links	Based on inbound links and outbound links	Based on different interaction metrics (e.g. votes, ratings, saves, comments, trackbacks, entries, inbound links etc.), which are gained from various services on the internet combined with PLEM data
Computation of the rank value	Query independent: The PageRank values of the pages on the Web are computed and saved off-line	Query dependent: The rank values of the pages are computed at the query time	Query independent: The rank values of the learning elements are computed on server backend and saved off-line
Time consideration	Does not consider time. Outdated contents and pages might still be ranked very high	Does not consider time	Does not consider time
Prominence consideration	The prominence or importance of pages that cast the vote is considered	The importance of pages that cast the vote is considered	Democracy, all votes are considered equal
Number of processed documents	It processes all relevant pages	It is processed on a small subset of relevant pages. Only $t$ (typically set to about 200) highest ranked pages, which assume to be highly relevant to the search query, are processed	It processes all relevant pages
Scores	Authority	Hub and Authority	Authority
Mashup	-	-	CoCoRank can be viewed as a mashup of ranking services
Algorithm	$P(i) = (1 - d) + d \sum_{(j,i) \in E} \frac{P(j)}{O_j}$	$a(i) = \sum_{(j,i) \in E} h(j)$ $h(i) = \sum_{(i,j) \in E} a(j)$	$r(i) = \sum_{i \in E, j \in S} p_j \cdot v_j$

	where $P(i)$ = PageRank score of page $i$ $O_j$ = number of out-links of page $j$ $d$ = dumping factor	where $a(i)$ = authority score of page $i$ $h(i)$ = hub score of page $i$	where $r(i)$ = rank value of learning entity $i$ $p_j$ = prestige of service $j$ $v_j$ = number of votes retrieved from service $j$
--	---	---	--

## 7. Conclusion

In this paper, we addressed the importance of **self-organized learning** within increasingly complex and fast changing learning environments. We discussed the concept of **Personal Learning Environment**, which offers a learner-centric view of learning, that takes a small pieces, loosely joined approach, characterized by the freeform use of a set of learner-controlled tools, the bottom-up creation of knowledge ecologies, and a shift from knowledge-push to **knowledge-pull**. We followed by a discussion of the problem of **knowledge overload**, as a consequence of a knowledge-pull approach to learning, and highlighted how the **collective intelligence**, when the whole is greater than the sum of its parts, can help learners cope with the problem of knowledge overload, based on the **Long Tail theory**. We then presented the design and implementation details of **PLEM**, a Web 2.0 driven service for personal learning management that acts as a **Long Tail aggregator** and **filter** for learning. The primary aim of PLEM is to tap the wisdom of crowds to help learners find quality in the Long Tail.

## References

- Anderson, C. (2006) The Long Tail: Why the Future of Business is Selling Less of More. Hyperion.
- Argyris, C. (1991) Teaching Smart People How to Learn. *Harward Business Review*. HBS Press.
- Argyris, C., Schön, D. A. (1978) Organizational Learning, A Theory of Action Perspective. Reading, Massachusetts: Addison-Wesley.
- Argyris, C., Schön, D. A. (1996) Organizational Learning II: Theory, Method and Practice. Reading, Massachusetts: Addison-Wesley.
- Brown, J. S. & Adler, R. P. (2008). Minds on Fire: Open Education, the Long Tail, and Learning 2.0. *EDUCAUSE Review*, 43 (1), 16–32.
- Chatti, M.A. & Jarke, M. (2009) Social Software for Bottom-up Knowledge Networking and Community Building M. D. Lytras, R. Tennyson and P. Ordóñez de Pablos (Eds.): Knowledge Networks: The Social Software

Perspective, 17-27. IDEA Group Publishing, Hershey, PA, USA.

Downes, S. (2005) E-Learning 2.0. ACM eLearn Magazine [On-line]. Available: <http://www.elearnmag.org/subpage.cfm?section=articles&article=29-1>

Google (2009). Google Web Toolkit, Google [On-line]. Available: <http://code.google.com/webtoolkit/>

Gordon, D. M. Control without hierarchy. *Nature*. 4468, 143.

Hase, S. & Kenyon, C. (2000) From Andragogy to Heutagogy. *ultibase Journal* [On-line]. Available: <http://ultibase.rmit.edu.au/Articles/dec00/hase2.htm>

Heylighen, F. (2002) Complexity and Information Overload in Society: why increasing efficiency leads to decreasing control [On-line]. Available: <http://pespmc1.vub.ac.be/Papers/Info-Overload.pdf>

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*. 46 (5), 604–632.

Levy, P. (1997) *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Plenum, New York.

Liu, B. (2006) *Web Data Mining*. Springer.

Miller, P. (2007) Swarm Theory. *National Geographic Magazine* [On-line]. Available: <http://ngm.nationalgeographic.com/2007/07/swarms/miller-text>

O'Reilly, T. (2007) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *International Journal of Digital Economics*. 65, 17-37.

Brin S. & Page L. (1998) The Anatomy of a Large-scale Hypertextual Web Search Engine. *Computer Networks*. 30 (1-7), 107-117.

Surowiecki, J. (2004) *The wisdom of crowds*. New York: Doubleday.

van Harmelen, M. (2006). *Personal Learning Environments*. [On-line]. Available: [http://octette.cs.man.ac.uk/jitt/index.php/Personal\\_Learning\\_Environments](http://octette.cs.man.ac.uk/jitt/index.php/Personal_Learning_Environments).

Figure 1. The Long Tail Graph (Anderson, 2006)

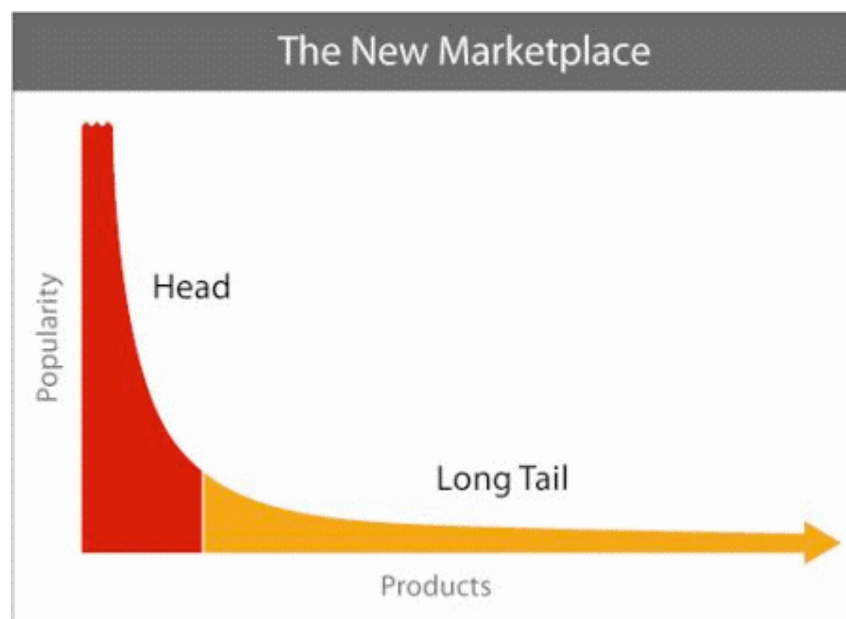


Figure 2. Abstract View of PLEM

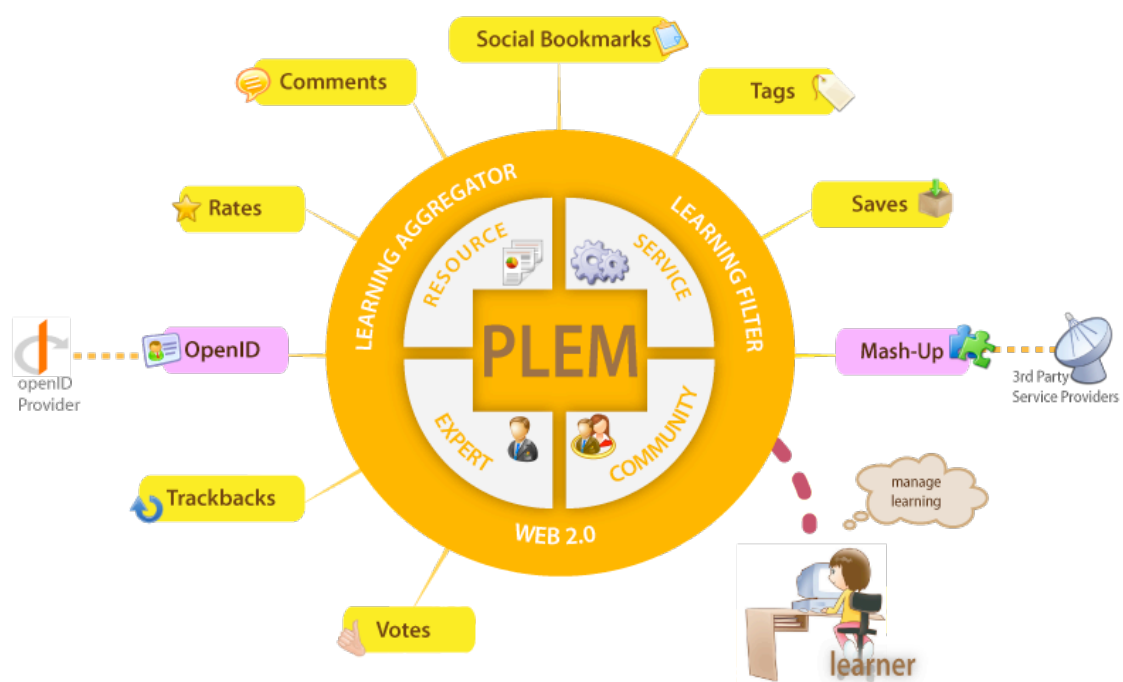


Figure 3. PLEM Architecture

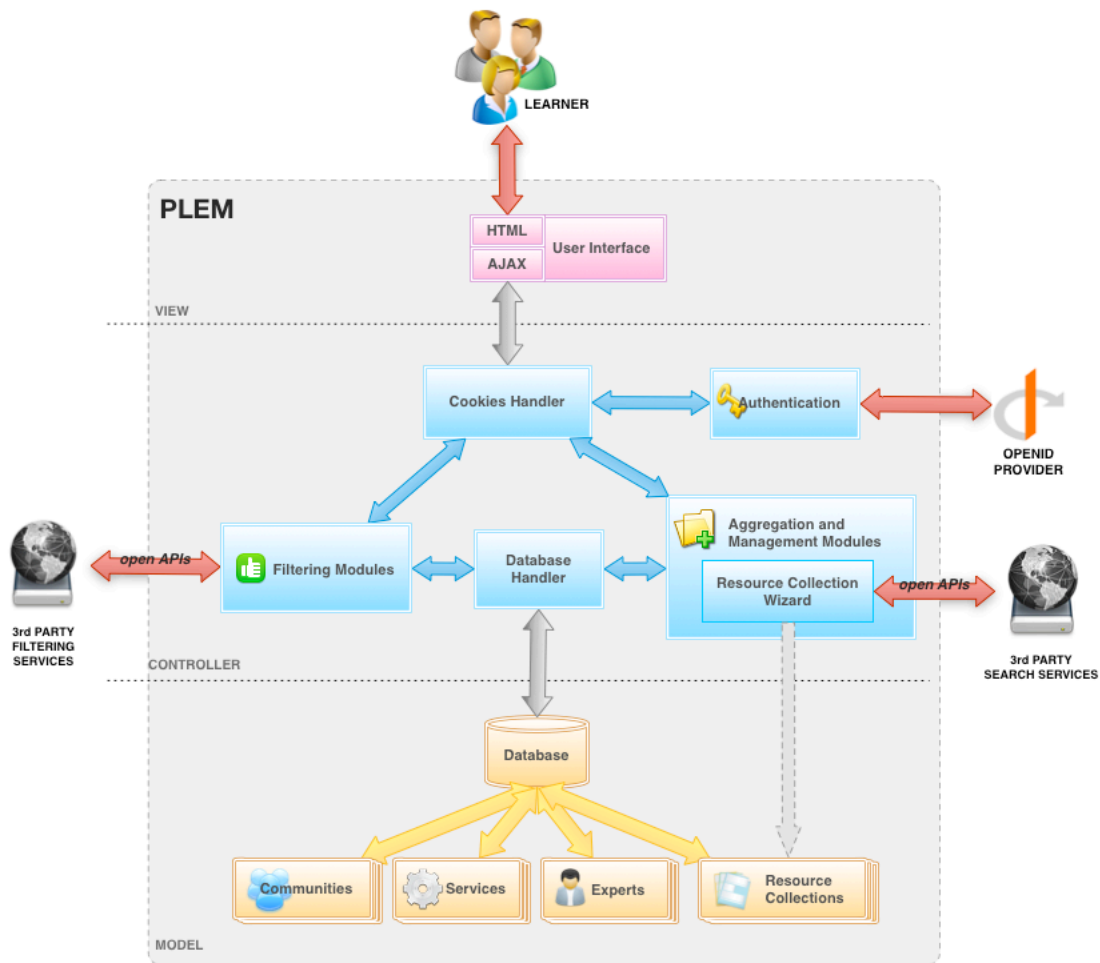


Figure 4. PLEM User Interface

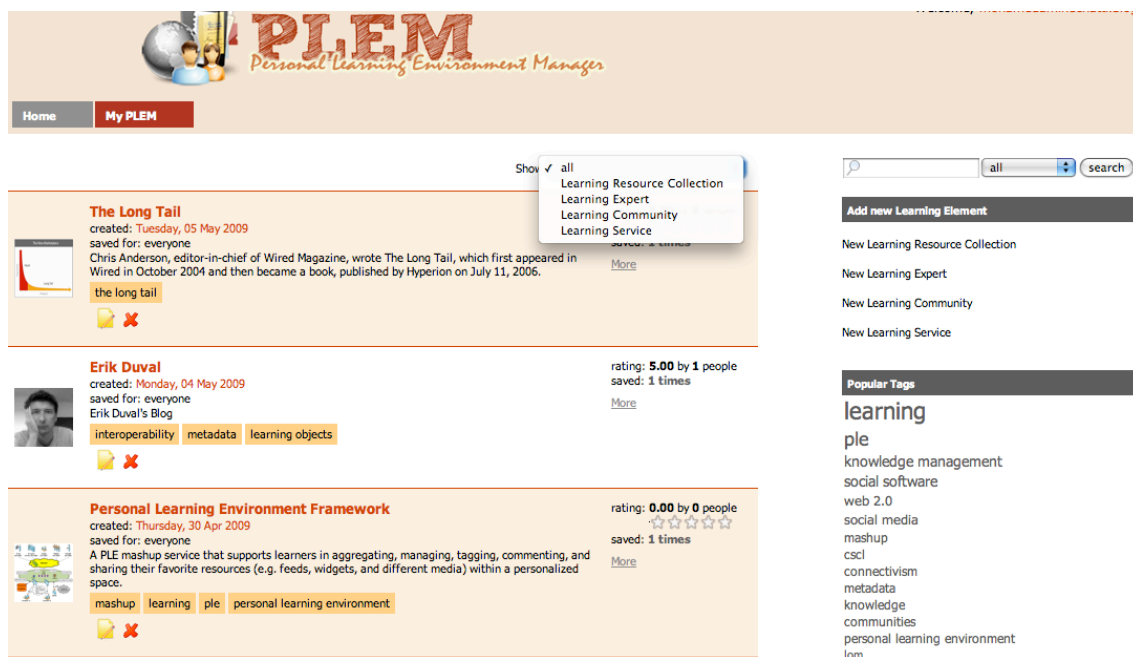


Figure 5. Federated Search in PLEM

Search Resources

Open Courses Blog **Video** Book Image Presentation

**Long Tail** **search!**

☐ Youtube ☐ Google Video

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



**Identifying "The Long Tail" - Chris Anderson**  
 Tuesday, 05 Dec 2006  
 Complete video at: [fora.tv](http://fora.tv) "Wired" editor and author Chris Anderson explains his theory of "The Long Tail," using the music and film industries as examples. ----- You know something is up when an audience member is taking cell phone photos of the presenter's slides for instant transmittal to a business partner. Chris Anderson does have killer slides, full of exuberant detail, defining the exact shape of the still emerging opportunity space for finding and selling formerly infindable and ...  
[+ ADD TO CART](#)

Resources Cart

- Identifying "The Long Tail" - Chris Anderson
- The Long tail
- The Long Tail: Why the Future of Business Is Selling Less of More
- long\_tail\_graph.gif

Clear cart

Add Resource Manually

Create Learning Resource Collection

Update Existing Learning Resource Collection

Figure 6. An example of an aggregated learning resource collection in PLEM



**The Long Tail**  
 first created on: Tuesday, 05 May 2009  
 by: [mohamedaminechatti.blogspot.com](http://mohamedaminechatti.blogspot.com)  
 Chris Anderson, editor-in-chief of Wired Magazine, wrote The Long Tail, which first appeared in Wired in October 2004 and then became a book, published by Hyperion on July 11, 2006.  
[the long tail](#)

Resources



**long\_tail\_graph.gif**  
 What Is The Long Tail?



**The Long Tail: Why the Future of Business Is Selling Less of More**  
 Book ID : ISBN1401302378  
 Published in : 2006  
 Number of pages : 256



**The Long Tail**  
 Chris Anderson's Blog



**Identifying "The Long Tail" - Chris Anderson**  
 Complete video at: <http://fora.tv/fora/showthread.php?t=453> "Wired" editor and author Chris Anderson explains his theory of "The Long Tail ...



**Minds on Fire: Open Education, the Long Tail, and Learning 2.0**  
 John Seely Brown and Richard P. Adler. EDUCAUSE Review, vol. 43, no. 1 (January/February 2008): 16-32

Figure 7. Ranking of Learning Entities in PLEM

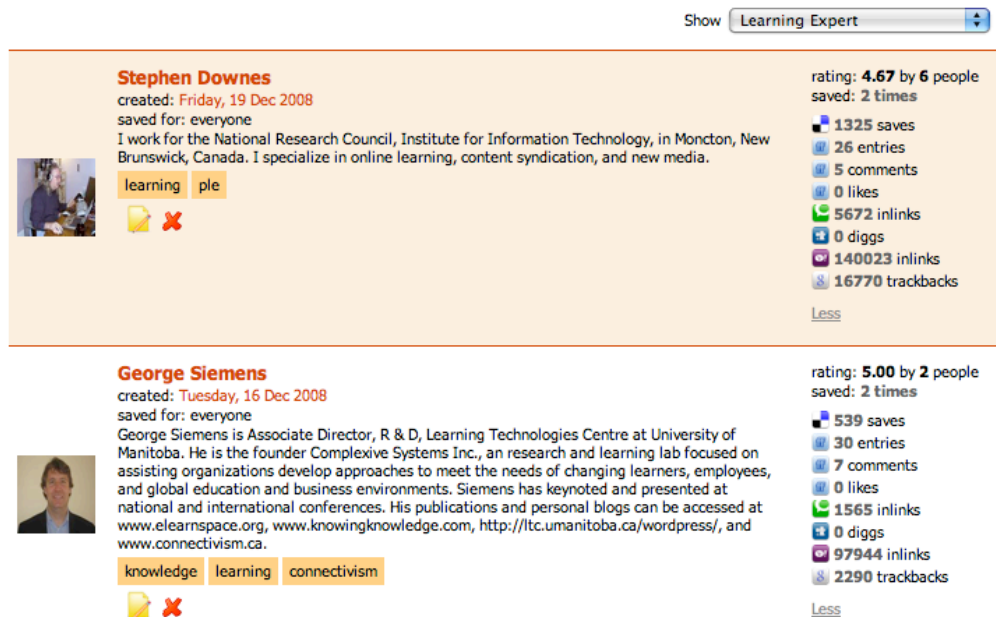


Figure 8. Flow Chart of the CoCoRank Algorithm

